2.0 DEFINITION AND MEASUREMENT OF METHOD VARIABILITY IN WET TESTING

The terms used to express toxicity test results are defined in this chapter, and methods for quantifying WET test method variability are discussed. Additional terms used throughout this document, along with their definitions, are provided in the Glossary as part of the front matter of this document.

2.1 Terms and Definitions

Biological endpoints are the biological observations recorded when conducting toxicity tests. These observations may include the number of surviving organisms or the number of young produced. There are two basic types of biological endpoints: responses recorded as response/no response (e.g., dead or alive) are quantal data; responses recorded as a measured response (e.g., weight) or as a count (e.g., number of young produced) are considered continuous data. For most WET tests, the observations for each tested concentration are combined and then reported as an average or percentage to represent the biological endpoint. For example, the fathead minnow larval survival and growth chronic test method has two biological endpoints (i.e., percent survival and average dry weight for each test concentration).

Effect concentrations are concentrations of a test material (i.e., effluent, referent toxicant, receiving water) derived from the observed biological endpoints followed by data analysis using either hypothesis testing procedures or point estimate techniques. Effect concentrations derived using point estimation techniques represent the concentration of a test material at which a predetermined level of effect occurs. For example, LC50 is the lethal concentration at which 50 percent of the organisms respond. Effect concentrations commonly estimated for WET methods are LC50, EC50 (effect concentration at which a 50-percent effect occurs), and IC25 (inhibition concentration at which a 25-percent effect occurs). Hypothesis test methods are used to determine the no observed effect concentration (NOEC). The NOEC represents the highest effect concentration in the test concentration response that is not significantly different from the control response. Multiple statistical endpoints can be derived for each WET method. For example, the endpoints for the fathead minnow larval survival and growth chronic test can be reported as an EC25 for growth, an NOEC for growth, an LC50 (or EC50) for survival, and an NOEC for survival.

2.2 Defining WET Test Variability

As with any measurement process, WET tests have a degree of variability associated with the test method performance. Three measures of variability related to WET tests are within-test variability, within-laboratory variability, and between-laboratory variability.

- **Within-test (intra-test) variability** is the variability in test organism response within a concentration averaged across all concentrations of the test material in a single test.

- **Within-laboratory (intra-laboratory) variability** is the variability that is measured when tests are conducted using specific methods under reasonably constant conditions in the same laboratory. Within-laboratory variability, as used in this document, includes within-test variability. The American Society for Testing and Materials (ASTM) uses the term “repeatability” to describe within-laboratory variability. Repeatability is estimated (as a sample variance or standard deviation) by repeating a test method under realistically constant conditions within a single laboratory.

- **Between-laboratory (inter-laboratory) variability** is the variability between laboratories. It is measured by obtaining results from different laboratories using the same test method and the same

test material (e.g., reference toxicant). Between-laboratory variability, as used in this document, does *not* include the within-laboratory component of variance. ASTM uses the term "reproducibility" to describe between-laboratory variability. Reproducibility is estimated by having nearly identical test samples (duplicates or splits) analyzed by multiple laboratories using similar standard methods. Although reproducibility is generally synonymous with between-laboratory variability, estimates of reproducibility may combine within-laboratory and between-laboratory components of variance, making between-laboratory variability numerically larger than within-laboratory variability as defined above.

For purposes of consistency, EPA uses the terms within-laboratory and between-laboratory variability throughout this document.

Numerous factors can affect the variability of any toxicity test method. These factors include the number of test organisms, the number of treatment replicates, randomization techniques, the source and health of the test organisms, the type of food used, laboratory environmental conditions, and dilution water quality. The experience of the analyst performing the test, analyzing the data, and interpreting the results may also affect variability (Grothe et al. 1996, Fulk 1996).

2.3 Quantifying WET Test Variability

Historically, information on the variability of toxicity tests has been developed using effect concentrations, such as the NOEC, EC25, EC50, and LC50 for survival, fecundity, and growth. Variability measures should be quantified based on the end use of the data (i.e., effect concentrations) and be directly related to the WET permit requirement. Typically, the effect concentrations are the endpoints used for evaluating self-monitoring results. The variability of the effect concentrations is quantified by obtaining multiple test results under similar test conditions using the same test material. For example, the sample standard deviation and mean for EC25 obtained from multiple monthly reference toxicant tests for the fathead minnow survival and growth chronic test conducted at one laboratory would quantify "within-laboratory" variability for that laboratory. EPA used this approach to evaluate data for the development of this document (see Chapter 3).

Examining variability for each effect concentration of each biological endpoint for each test method is essential. The biological endpoints may be different for various toxicants and effluents. One biological endpoint, such as reproduction, may be more sensitive to a certain toxicant than another endpoint, such as survival. That sensitivity may be reversed for a different toxicant. Alternatively, an endpoint may be more sensitive to one toxicant than another toxicant.

Three other measures of variability (which are not addressed in this document) that have been applied to WET tests are:

1. Determine the variability of the biological endpoint response. For example, the variance of the biological response (e.g., growth and survival) can be calculated. This approach is useful, but does not quantify variability of the WET test effect concentration, which is important in the context of this document.

2. Quantify the uncertainty of each test point estimate (e.g., the EC50, EC25, or LC50) using confidence intervals, which reflect within-test variability.

3. Use the standard deviation to quantify the uncertainty in the mean of the replicate response at each concentration within a particular test. For example, laboratories can compare the standard deviations of the average weight of fathead minnow larvae in four chronic tests at one test concentration, such as 1 mg/L sodium chloride. These standard deviations may be pooled across

all the concentrations when data have been transformed (if necessary) to give similar variances at each concentration. From the pooled variance, one may calculate a minimum significant difference (MSD) value, which is a useful indication of test sensitivity (see Chapters 3 and 5). In this document, the standard deviation at each concentration was not evaluated as a measure of variability. However, the MSD was considered as a measure of WET test variability.

This page intentionally left blank.